SAO Special Report No. 122

COMBINATIONS OF LEAST-SQUARES APPROXIMATIONS IN THE CASE OF CORRELATED VARIABLES

by

Pravin L. Kadakia

Smithsonian Institution
Astrophysical Observatory

Cambridge 38, Massachusetts

COMBINATIONS OF LEAST-SQUARES APPROXIMATIONS IN THE CASE OF CORRELATED VARIABLES[1, 2]

Pravin L. Kadakia[3]

When fitting an equation or polynomial curve of some degree to a large number of observations by the method of least squares (Linnik, 1961), we are often faced with the problem of too few storage locations in the memory of a large-scale digital computer such as the IBM-7090 or any other series. Sometimes we are also faced with the problem of estimating a mean vector of n independent sample mean vectors of the same population or the same physical quantity with their corresponding n variance-covariance matrices (or weight matrices) for which off-diagonal elements are nonzero. The following example illustrates the latter case.

At this Observatory we have a Differential Orbit Improvement program (DOI). At one stage it computes the correction parameters ($\Delta x, \Delta y, \Delta z$) to the geodetic station coordinates. These corrections are correlated among themselves. For M independent runs of the program with different sets of observations for the same station coordinates, we have a set of vectors (Δx_i, Δy_i, Δz_i), i = 1,2 ... , M, each representing

[1] The complete program has been written up in FAP language for IBM-7090 under the title "Super Least-Squares Program," and is in use at this Observatory.

[2] This work was supported by Grant NsG 87-60 from the National Aeronautics and Space Administration.

[3] Smithsonian Astrophysical Observatory, Cambridge, Mass.

a correction to the coordinates, with variance-covariance matrices

$$\sigma^k_{i,j} = \begin{pmatrix} \sigma^k_{11} & \sigma^k_{12} & \sigma^k_{13} \\ & \sigma^k_{22} & \sigma^k_{23} \\ & & \sigma^k_{33} \end{pmatrix}$$

and standard deviation σ_k ($k = 1,2, \ldots, M$). The problem is to find the average correction vector of different sample correction vectors from the information of sample variance-covariance matrices for which the off-diagonal elements are nonzero.

This paper discusses a method based on statistical estimation theory.

Statement of the problem

Let $X_1, X_2, \ldots, X_n$ be N random variables and distributed according to multivariate normal distribution (Anderson, 1958)(Gaussian distribution or Error distribution). Associated with these N random variables is a mean vector $(\mu_1, \mu_2, \ldots, \mu_N)$ whose μ_N element is the population mean of the X_N variable and a variance-covariance matrix, $\Sigma_{(NxN)}$. Also let $(X_1, X_2, \ldots, X_N)$ be a randomly correlated vector (correlation may be high or very low). Given $M(\geq 30)$ sample observations such as $x_{i,1}, x_{i,2}, \ldots, x_{i,N}$ ($i = 1,2, \ldots, M$) for each variable from an N-variate normal population with known covariance matrices $\sigma^k_{i,j}$ and standard deviations σ_k $(k = 1,2, \ldots , M)$, <u>we ask for the estimates of the mean vector</u> $(\mu_1, \mu_2, \ldots , \mu_N)$, <u>average covariance matrix Σ of covariances matrices</u> $\sigma^k_{i,j}$, <u>and average standard deviation of</u> σ_k <u>$(k = 1,2, \ldots , M)$</u>.

Estimation theory

Let $\hat{\mu} = (\hat{\mu}_1, \hat{\mu}_2, \ldots , \hat{\mu}_N)$ and $\hat{\Sigma}$ be the estimates of $\mu = (\mu_1, \mu_2, \ldots , \mu_N)$ and Σ, respectively. The problem is to estimate $\hat{\mu}$ and $\hat{\Sigma}$, or, more precisely, to find a mathematical formula for $\hat{\mu}$ and $\hat{\Sigma}$.

The random vector $(X_1, X_2, \ldots , X_N)$ has a multivariate normal distribution, and the joint multivariate density function (Feller, 1950; Hoel, 1954; Mood, 1950) of this random vector $(X_1, X_2, \ldots , X_N)$ is given by the expression:

$$c \exp \left\{ - \frac{1}{2} (X-\mu)' \Sigma^{-1} (X-\mu) \right\}, \text{ where } c \text{ is a constant;}$$

$$(X-\mu) \text{ is a column vector } = \begin{matrix} X_1 - \mu_1 \\ X_2 - \mu_2 \\ \vdots \\ X_N - \mu_N \end{matrix}$$

and $(X-\mu)'$ is the transpose of $(X-\mu)$.

We have M sample observations on the random vector with their corresponding variance-covariance matrices, and their likelihood function L (Anderson, 1958; Hoel, 1954) is given by

$$L = \frac{\exp\left[-\frac{1}{2} \sum_{k=1}^{M} (X_k - \mu)' \left[\sigma^k_{i,j} \right]^{-1} (X_k - \mu) \right]}{\prod_{k=1}^{M} \left\{ \left| \sigma^k_{i,j} \right| (2\pi)^N \right\}^{\frac{1}{2}}}$$

where

$$(X_k - \mu) = \begin{bmatrix} x_{k,1} & -\mu_1 \\ x_{k,2} & -\mu_2 \\ \vdots & \\ x_{k,N} & -\mu_N \end{bmatrix}$$

and $\left[\sigma^k_{i,j}\right]^{-1}$ is the inverse matrix of the variance-covariance matrix of the k^{th} sample observation.

The variance-covariance matrix of the k^{th} observation $\left[\sigma^k_{i,j}\right] = \left[A^k_{i,j}\right]^{-1} \sigma^2_k$, where $\left[A^k_{i,j}\right]$ is the normal matrix of normal equations that are obtained by using the method of least squares, and σ_k is the standard deviation of the residuals for the k^{th} observation.

Hence

$$\left[\sigma^k_{i,j}\right]^{-1} = \left\{\left[A^k_{i,j}\right]^{-1}\right\}^{-1} \Big/ \sigma^2_k = \left[A^k_{i,j}\right] \Big/ \sigma^2_k \, ,$$

and the likelihood function

$$L = \frac{\exp\left\{-\frac{1}{2}\sum_{k=1}^{M}(X_k-\mu)'\frac{\left[A^k_{i,j}\right]}{\sigma_k^2}(X_k-\mu)\right\}}{\prod_{k=1}^{M}\left\{|A^k_{i,j}|\sigma_k^2(2\pi)^N\right\}^{-1}}$$

$$= \frac{\exp\left[-\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{M}\frac{A^k_{i,j}}{\sigma_k^2}\left(x_{ki}-\mu_i\right)\left(x_{kj}-\mu_j\right)\right]}{\prod_{k=1}^{M}\left\{|A^k_{i,j}|\sigma_k^2\ (2\pi)^N\right\}^{-1}} .$$

By using the maximum-likelihood estimation method (Mood, 1950; Fraser, 1958) which maximizes the likelihood function for the given sample observations (i.e., the maximum-likelihood estimate of the parameter is the point at which the likelihood function has a maximum), we can find the estimates $\hat{\mu}$ and $\hat{\Sigma}$.[1] In the likelihood function the value of vector $(X_1, X_2, \ldots, X_N)$ is fixed at the sample values and L is a function of μ and Σ.

To emphasize that the quantities (μ and Σ) are variables and not parameters we replace them by $\hat{\mu}$ and $\hat{\Sigma}$. Then the logarithm of the likelihood function is

$$\log L = -\log\left[\prod_{k=1}^{M}\left(|A^k_{i,j}|\sigma_k^2\ (2\pi)^N\right)^{\frac{1}{2}}\right]$$

[1] The maximum likelihood method and the least-squares method give the identical estimate for the parameter provided the distribution of the residuals is normal.

$$-\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{M}\left[\frac{A_{i,j}^{k}}{\sigma_k^2}\left(x_{k,i}-\hat{\mu}_i\right)\left(x_{k,j}-\hat{\mu}_j\right)\right].$$

Since log L is an increasing function of L, its maximum is at the same point in the space of $\hat{\mu}$ and $\hat{\Sigma}$ as the maximum of L.

To find the estimate, we compute

$$\frac{\partial \log L}{\partial \hat{\mu}_i} = \sum_{j=1}^{N}\sum_{k=1}^{M}\left[\frac{A_{i,j}^{k}}{\sigma_k^2}\left(x_{k,j}-\hat{\mu}_j\right)\right] = 0, \qquad i = 1,2, \dots , N.$$

Hence we obtain

$$\sum_{j=1}^{N}\sum_{k=1}^{M}\left[\frac{A_{i,j}^{k}}{\sigma_k^2}\right]\hat{\mu}_j = \sum_{j=1}^{N}\sum_{k=1}^{M}\left[\frac{A_{i,j}^{k}}{\sigma_k^2}\, x_{k,j}\right] \quad i = 1,2, \dots , N,$$

where $\hat{\mu}_j$ $(j = 1,2, \dots , N)$ are N unknown coefficients, i.e., the estimate of μ is

$$\hat{\mu} = \begin{bmatrix} \hat{\mu}_1 \\ \hat{\mu}_2 \\ \cdot \\ \cdot \\ \cdot \\ \hat{\mu}_N \end{bmatrix}$$

which is the mean vector of different sample observations. Finally, we have N equations and N unknown coefficients.

In matrix notation: $AZ = B$, where Z is the solution vector.

$$Z = \begin{bmatrix} \hat{\mu}_1 \\ \hat{\mu}_2 \\ \vdots \\ \hat{\mu}_N \end{bmatrix}, \qquad A = \begin{bmatrix} \sum_{k=1}^{M} \left(\frac{A_{11}^k}{\sigma_k^2}\right) & \sum_{k=1}^{M} \left(\frac{A_{12}^k}{\sigma_k^2}\right) & \cdots & \sum_{k=1}^{M} \left(\frac{A_{1N}^k}{\sigma_k^2}\right) \\ - & \sum_{k=1}^{M} \left(\frac{A_{22}^k}{\sigma_k^2}\right) & \cdots & \sum_{k=1}^{M} \left(\frac{A_{2N}^k}{\sigma_k^2}\right) \\ & \vdots & & \\ - & - & & \sum_{k=1}^{M} \left(\frac{A_{NN}^k}{\sigma_k^2}\right) \end{bmatrix}$$

$$B = \begin{bmatrix} \sum_{j=1}^{N} \sum_{k=1}^{M} \left(\frac{A_{1j}^k \; x_{kj}}{\sigma_k^2}\right) \\ \sum_{j=1}^{N} \sum_{k=1}^{M} \left(\frac{A_{2j}^k \; x_{kj}}{\sigma_k^2}\right) \\ \sum_{j=1}^{N} \sum_{k=1}^{M} \left(\frac{A_{Nj}^k \; x_{kj}}{\sigma_k^2}\right) \end{bmatrix}.$$

Z yields the mean vector of different sample vectors using the information of variance-covariance matrices for which off-diagonal elements are non-zero, and their corresponding average variance-covariance, $\hat{\Sigma}$, is the multiple of the inverse matrix of A and σ^2, where

$$\sigma^2 = \frac{\sum_{k=1}^{M} (N_k-1)\ \sigma_k^2}{\sum_{k=1}^{M} N_k - M} ,$$

and N_k is the number of observations used to compute σ_k, i.e.,

$$\hat{\Sigma} = A^{-1}\ \sigma^2 .$$

Application

Consider the first problem mentioned in the first paragraph of the introduction. Suppose we have 1800 pairs of (y, x) observations collected over a 30-period time interval and we desire to fit the second degree polynomial curve $y = a + bx + cx^2$ to the 1800 observations of the 30-period time interval.

Assume further that for each period there are 60 (y, x) observations and that the curve $y = a + bx + cx^2$ has been fitted to these 60 (y, x) observations by using the least-squares method. For each period, therefore, we have coefficients vector (a,b,c) the corresponding normal matrix, and the corresponding standard deviation of the residuals.

As this was done for each of the 30 periods we will have (a_i, b_i, c_i) with the corresponding normal matrix A_i (or variance-covariance matrix which is equal to $A_i^{-1}\ \sigma_i^2$) and standard deviation σ_i ($i = 1,2, \ldots, M$, with $M = 30$ for this case).

In order to estimate the (a,b,c) and the corresponding variance-covariance matrix for the entire 30-period interval without having to deal with the 1800 (y,x) observations collectively, compute A, B as defined earlier in this paper, then solve $AZ = B$, i.e. solve 3 equations and 3 unknowns.

The method developed in this paper may have manifold applications in various fields.

Acknowledgements

I appreciate the interest and help I received from Mr. M. Gaposchkin, Mr. I. Izsak and Dr. G. Veis, of the Smithsonian Astrophysical Observatory. I also wish to thank Professor William G. Cochran, of Harvard University, for his helpful comments.

The results in this paper are not claimed to be new since they are presumably known to many specialists who use the method of least squares in their work. But I have been unable to give a reference, and my results are presented here in view of their practical utility in the field of astronomy.

References

ANDERSON, T. W.
1958. An introduction to multivariate statistical analysis. John Wiley & Sons, New York.

FELLER, W.
1950. An introduction to probability theory and its applications, vol. 1. John Wiley & Sons, New York.

FRASER, D. A. S.
1958. Statistics, an introduction. John Wiley & Sons, New York.

HOEL, P. G.
1954. Introduction to mathematical statistics, 2nd ed. John Wiley & Sons, New York.

LINNIK, Y. V.
1961. Method of least squares and principles of the theory of observations. Pergamon Press, New York.

MOOD, A. M.
1950. Introduction to the theory of statistics. McGraw-Hill Book Co., New York.

NOTICE

This series of Special Reports was instituted under the supervision of Dr. F. L. Whipple, Director of the Astrophysical Observatory of the Smithsonian Institution, shortly after the launching of the first artificial earth satellite on October 4, 1957. Contributions come from the Staff of the Observatory. First issued to ensure the immediate dissemination of data for satellite tracking, the Reports have continued to provide a rapid distribution of catalogues of satellite observations, orbital information, and preliminary results of data analyses prior to formal publication in the appropriate journals.

Edited and produced under the supervision of Mr. E. N. Hayes, the Reports are indexed by the Science and Technology Division of the Library of Congress, and are regularly distributed to all institutions participating in the U.S. space research program and to individual scientists who request them from the Administrative Officer, Technical Information, Smithsonian Astrophysical Observatory, Cambridge 38, Massachusetts.